



Contribution ID: 142

Type: **Presentation**

Towards understanding identification, selection and appraisal in contemporary digital preservation practice

Tuesday 14 October 2025 12:36 (11 minutes)

Identification, selection and appraisal are key digital preservation activities when ingesting data objects for long-term preservation. This paper describes the approach taken to a major new global survey by the EOSC EDEN project, designed to improve understanding of current practices in this area, and the frameworks, standards and guidelines that support digital preservation professionals when tackling these challenges. We will situate the survey work in its wider context, outline our approach to question-building and analysis, and share what we hope our findings will tell us, including an overview of analysis and findings to date.

Regardless of which lifecycle model, conceptual framework or workflow is used by a digital preservation organisation, there is always a point at which data objects officially enter the digital preservation environment. However, these frameworks and models often don't clearly outline how, when and what criteria the decision to preserve that data is based upon. In addition, the term used to describe this decision process may be dependent on the domain in which the archive is embedded. Three terms commonly used for this process are: identification, selection, and appraisal.

These practices are key elements of digital preservation practice for any size of organisation, particularly when the data objects are intended for long-term data preservation, with the funding commitment and responsibility that implies. But how do on-the-job digital preservation professionals approach these key activities? Which standards, guidelines and frameworks do they refer to? What levels of quality are measured and how? Can these quality metrics be used for re-appraisal processes along the lifecycle if data is not to be stored for the long-term? And what does "long-term" mean anyway?

The newly-initiated European Open Science Cloud (EOSC) project, 'Enhancing Digital preservation strategies at European and National level'(EDEN) is a three-year (2025-2027) research initiative funded by the EU Horizon Europe programme, that aims to tackle these (among other) questions. The project targets research data archives and addresses the questions of how data quality for digital preservation can be defined, and how this definition can be used in decision-making processes for ongoing preservation. A first step within this work is a global survey to understand how the community currently conducts these decision-making processes for digital preservation.

The questioning approach is as follows. We will ask the participants about:

- Descriptive information from each participant.
- Familiarity with frameworks and guidelines for appraisal, identification and selection of data for long-term preservation: a list of well-established frameworks and guidelines are presented. The respondent indicates how familiar they are with each and can list any further reference or guidance resources that they use.
- Definition of "long-term" preservation: the respondent is asked whether their organisation has a working or agreed definition of this and if so, the source of this definition. If they are working without an agreed definition, the respondent is invited to report why this is the case.
- Pathways of data into the archive: the respondent is asked how digital objects enter the organisation for long-term preservation. Choices include self-deposit on a voluntary basis, mandated deposit such as legal or funder mandated deposit, proactive collection building such as harvesting or classical collection building in libraries, and/or third-party data preservation on a contractual level.
- Practices in assessment of quality of data objects at ingest: the respondent is offered the choice of various aspects of data objects (technical quality, content/information quality, quality of technical metadata, quality of descriptive metadata, quality of administrative metadata). For each aspect, the respondent can specify the types of quality assessment undertaken. There is a further question on the frequency of reassessment.
- How long digital objects are initially preserved.

- What happens to digital objects after the agreed initial preservation period.
- Discipline-specific questions submitted by EDEN WP3 on requirements including metadata standards used, file formats preferred, handling of sensitive data, risks, and community needs, all with a discipline-specific lens.

Messaging about and within the survey is intended to avoid discipline-specific language or unnecessary technical terms. Also, our survey is designed to be used across the global research, cultural heritage, and industry sectors, so it was also necessary to ensure that our language is as sector-neutral and globally appropriate as possible.

In addition, the survey is designed for all those working within the digital preservation organisational context, whether at senior management, middle management, or practitioner levels. Our choice of language aims to accommodate these various staff levels.

By this work, we hope to better understand:

- Which quality aspects are currently checked (and if there's a stronger leaning towards content quality or towards technical quality depending on institution type);
- If there is a shared definition of 'long-term';
- How the organisations who define 'long-term' as a shorter period (e.g. less than 10 years) approach quality checks, and succession planning for these data objects.

The EOSC EDEN project will support the digital preservation community and contribute to the European Open Science Cloud (EOSC) through efforts to better understand current digital preservation practice and to provide appropriate guidance and resources to the community. This includes the survey described here, designed to better understand current identification, appraisal, and selection practices when dealing with the ingest of data objects, and the reference materials that digital preservation professionals currently use to guide them.

Primary authors: Dr MOLLOY, Laura (CODATA); Mx LINDLAR, Micky (Leibniz Information Centre for Science and Technology)

Presenters: Dr MOLLOY, Laura (CODATA); Mx LINDLAR, Micky (Leibniz Information Centre for Science and Technology)

Session Classification: Presentations Session 4: Data Stewardship

Track Classification: SciDataCon Persistent Themes: Data Stewardship